AI Based Chatbot System With Long Term Memory
Shruti Katkade¹, Urjita Kerkar², Pravin Bhilare³, Manish Gupta⁴, Payel Thakur⁵
Assistant Professor³
Department of Computer Engineering
PCE, Navi Mumbai, India

Abstract:
This paper explores a specific deep learning technique to build a conversational agent. In recent times the popularity of chat-oriented dialogue systems is on rise as they attempt to get into daily life and achieve some commercial success. Previous approaches used simple keywords & pattern matching methodologies, answering in a static manner irrespective of previous conversions. As an improvement to this technology would be a system that will work with sequence to sequence framework. Our proposed model makes use of this framework. Given the previous sentence or sentences and next sentence in a conversation, model converses by predicting improvement to this technology would be a system that will work with sequence to sequence framework. Our proposed model makes use of this framework. Given the previous sentence or sentences and next sentence in a conversation, model converses by predicting the next sentence. The distinctive feature of our model is that it can be trained end-to-end hence requires much fewer hand-crafted rules. This straightforward model can generate simple conversations given a large conversational training dataset.

Keywords: Deep Learning, Recurrent neural network, end to end memory, LSTM model, seq-to-seq model.

I. INTRODUCTION

Before deep learning hit the scene a few years ago all chatbots used hard coded rules. The programmer had to think of all of the possible responses to a question and handcraft rules accordingly. Recently there has been a growing interest in developing of chatbots using deep learning that use end to end system as it is less cumbersome to build than the previous methodologies in which different components are trained to carry out a specific task and the result of each component is collectively used to predict a response. Although End to end systems are preferred when you have really large dataset available to train your model. Many researchers have explored the use of deep learnings recurrent neural network architecture to develop efficient chatbots. The most popular model used nowadays is the seq-to-seq recurrent neural network model used for sentence generation. Seq-to- Seq was initially proposed for machine translation, but it can be applied in generating the conversation as well. The seq2seq model has two units encoder unit and a decoder unit. The encoder unit takes the input sequence from the system and train on it then it passes the last state of its recurrent layer as an initial state to the first recurrent layer of the decoder part. The decoder takes the last state of encoders last recurrent layer and uses it as an initial state to its first recurrent layer, the sequence we want to get is the input of decoder. This type of seq2seq model has shown impressive performance in various domains like speech recognition, machine translation, question answers etc. Conversation generation using deep learning includes two types of models: Retrieval based model and generative model. Retrieval based model has a repository of responses and uses some kind of heuristics to pick an appropriate response based on input and content hence it cannot handle unseen cases. On the other hand generative models create a response from scratch. But they are more prone to grammatical errors which can be handled if these models are trained precisely. This research is on generative models based method LSTM for conversation generation trained end to end.

II. LITERATURE SURVEY

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Paper</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chatbot : Artificially Intelligent Conversational Agent, Ayesh Shaikh, Geetanjali Phalke, 2017</td>
<td>To design a chatbot system for a particular college using the divide and conquer strategy. It makes use of the Dialogue Acts Classifiers, POS-taggerData Mining, Pattern matching, Knowledge Database, cloud services to store responses.</td>
<td>Chatbot system is developed only for android users.</td>
</tr>
<tr>
<td>2</td>
<td>A Neural-network based chatbot, Milla T Mutiwokuziv, Melody W Chanda, 2017</td>
<td>To explore the capabilities of deep neural networks by designing a chatbot using recurrent neural networks (RNN). Implementing sequence to sequence long short-term memory cell neural network (LSTM) in conjunction with Google word2vec.</td>
<td>Textual input provided to the system. Voice recognitio and voice synthesis can be provided as input and output to system.</td>
</tr>
<tr>
<td>3</td>
<td>Long Short-Term Memory</td>
<td>To develop an intelligent ‘chatbot’, which generates</td>
<td>Current system is particularly...</td>
</tr>
</tbody>
</table>
Networks for Automatic Generation of Conversations, Tomohiro Fujita, 2017

Conversational sentences via recurrent neural network and its coupled memory unit, long short-term memory (LSTM). Vulnerable to the case of Notation Change and the sequential outputs are frequently neither linguistic nor contextual valid.

Chatbot Using Gated End-to-End Memory Networks, Mill a T Mutiwokusviza, Melody W Chanda, 2017

To explore the capabilities of deep neural networks by designing a chatbot using recurrent neural networks(RNN). Implementing sequence to sequence long short-term memory cell neural network (LSTM) in conjunction with Google word2vec. Textual input provided to the system. Voice recognition and voice synthesis can be provided as input and output to system.

Chatbot Using A Knowledge in Database, Bayu Setiaji, 2016

Bigram which divides input sentence as two letters of input sentence. Database employed as knowledge storage, interpreter has been employed as stored programs of function and procedure sets for pattern-matching.

A Neural Conversational Model, Oriol Vinyals, 2015

To present a simple approach for conversational modelling using the recently proposed sequence to sequence framework. Chatbot is trained end-to-end and thus requires much fewer hand-crafted rules. The model does not capture a consistent personality.


To design an example-based chat-oriented dialogue system with personalization framework using long-term memory. To reduce human labor, annotation cost and build rapport with the user. EBDM detecting important topic keywords automatically will enable system to extent to other language.

III. PROPOSED WORK

An LSTM(Long term short term memory) network is a recurrent neural network that has LSTM cells in place of neural network layers. They can remember information for a long period of time. The architecture specifies that our model has memory and is trained end to end, with more accuracy as Bidirectional LSTM has been used, which makes it stand out.

3.1. System Architecture

The system architecture is given in Figure 2. Each block is described in this Section.

3.2. Existing System

![Figure 1. LSTM model for conversation generation][2]

3.3. Proposed System

![Figure 2. Proposed system architecture](http://ijesc.org/)

### A. Block 1

The input to the system is sequence of sentences provided by user. This input string is split into words. Each of the word is indexed to form a word ID for the specific word. Given a word ID to the system, it retrieves the word
corresponding to the current id to find its next word ID. It then concatenates the two words for generating a candidate system snippet. Model’s internal state is kept updating until the candidate snippet output is terminated by symbol `<EOS>`.

**B. Block 2:** The embedding layers task is to perform the word embedding. Each word is represented as real valued vectors of hundreds of dimension in a predefined vector space. The vector is a distributed expression of each word.

**C. Block 3 and 6:** The tanh and ReLu are activation functions. To improve the representation keeping the expression size unaltered, the tanh and ReLu layer is introduced.

**D. Block 4 and 5:** The Bidirectional LSTM network is more powerful than LSTM. For any point in the sequence BLSTM stores entire sequential information about points before and after that point. In BLSTM, The forward and backward passes over the unfolded network over time are carried out in a similar way to regular network forward and backward passes, except that we need to unfold the hidden states for all time steps.

**E. Block 7:** The last block is the output layer. The generated response is again separated word by word and is provided with a word ID.

### 3. Requirement Analysis

The implementation details are given in this section.

#### Dataset and Parameters

**1. Movie Dialogue dataset:**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>movie_characters</td>
<td>9034</td>
</tr>
<tr>
<td>movie_conversations</td>
<td>9030</td>
</tr>
<tr>
<td>movie_lines</td>
<td>666256</td>
</tr>
<tr>
<td>movie_titles</td>
<td>616</td>
</tr>
</tbody>
</table>

**Software**

<table>
<thead>
<tr>
<th>SR NO.</th>
<th>NAME</th>
<th>SPECIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C.P.U with minimum clock speed:</td>
<td>Intel celeron 700 Mhz processor</td>
</tr>
<tr>
<td>2</td>
<td>Storage:</td>
<td>4GB of hard-drive space</td>
</tr>
<tr>
<td>3</td>
<td>R.A.M</td>
<td>4GB system memory</td>
</tr>
</tbody>
</table>

**Hardware**

<table>
<thead>
<tr>
<th>SR NO.</th>
<th>NAME</th>
<th>SPECIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Compatible Operating System:</td>
<td>with Ubuntu 14.04 or up.</td>
</tr>
<tr>
<td>2</td>
<td>Language:</td>
<td>Python, HTML, JavaScript, CSS.</td>
</tr>
<tr>
<td>3</td>
<td>Drivers:</td>
<td>Drivers, Soundcard drivers.</td>
</tr>
<tr>
<td>4</td>
<td>A.P.I:</td>
<td>Google Speech and Processing, Responsive voice JS, Spacy, Flask</td>
</tr>
</tbody>
</table>

### IV. ACKNOWLEDGMENT

It is our privilege to express our sincerest regards to our supervisor Prof. Payel Thakur for the valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of this work. We deeply express our sincere thanks to our Head of the Department Dr. Sharvari Govilkar and our Principal Dr. Sandeep M. Joshi for encouraging and allowing us to present this work.

### V. REFERENCES


[5]. https://www.cs.cornell.edu/~cristian/Cornell_Movie-Dial ogs_Corpus.html

[6]. https://medium.com/@BhashkarKunal/conversational-ai-chatbot-using-deep-learning-how-bi-Directional-lstm-machine-reading-38dc5cf5a5a3


[8]. www.esds.co.in